DOCUMENT RESUME

ED 113 375 TM 004 872

AUTHOR Woodley, Katheryn K.
TITLE Test-Wiseness: A Cognitive Function?
PUB DATE 75
NOTE 24p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Washington, D.C., March 31-April 2, 1975)
AVAILABLE FROM American College, Bryn Mawr, Pa. (Free)

EDRS PRICE MF-$0.76 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS *Adults; Anxiety; Certification; *Cognitive Ability; Essay Tests; Formative Evaluation; *Instructional Programs; Multiple Choice Tests; Occupational Tests; Program Effectiveness; Testing; *Test Wiseness

ABSTRACT

This paper reports the findings of an attempt to improve test-wiseness (TW) through direct instruction in selected test-taking strategies. TW was defined as "a cognitive function, subject to improvement through both general exposure to a wide variety of test items, and specific training in test-taking skills." The total investigation included: development and validation of a test instrument to measure TW; evaluation of the level of TW in an adult occupational group; program development and refinement; investigation of the psychological correlates of TW; and, formative and summative evaluations of a TW program. A program was developed to provide training in responding to multiple-choice and essay test items. The program combines instruction and measurement in a workbook format, with diagnostic testing and some branching. It was designed for use by adults, and focuses on strategies applicable to a wide variety of occupational or licensing examinations. Rules and strategies are emphasized, rather than practice in specific item types. Included are rules related to accuracy of item interpretation, deductive reasoning to eliminate absurd, similar or contradictory options, and cue-using strategies. A complex, multi-sample design was used for formative evaluations and additional evaluations will be completed. General findings to date have shown an increase in TW both on an internal and an external criterion and a decrease in test-related anxiety, after use of the program. Trends for increased consistency in test performance over time were noted. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED113375

TEST-WISENESS: A COGNITIVE FUNCTION?

PERMISSION TO REPRODUCE THIS [redacted] MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY

K. K. Woodley

TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE [redacted] OWNER."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

TM004 872

Katheryn K. Woodley
The American College, Bryn Mawr, Pa.
Paper presented at NCME
Washington, D. C.
April, 1975

Definition of Test-Wiseness

Test-wiseness (TW) is a term which most researchers have probably heard or used, and often without a true understanding of the meaning of this fairly specific term. As a behavior, it is often confused with guessing or risk-taking. As an explanation of test performance, it is often confused with bias or response sets, and very often is considered merely as part of undifferentiated error variance. To some people, the test-wise individual is seen as contributing to the unreliability of a test of knowledge, or interfering with the validity of a test of personality. In his analysis of sources of test variance, Stanley (1971) classified TW as one of the general and lasting characteristics of the individual. He maintained that while TW represents systematic variance, when unrelated to the criterion of interest, variations in the level of TW will serve to reduce the validity of the test. He considers test-wiseness as a real factor in almost any test score, since "freedom from emotional tension, shrewdness in guessing, and a keen eye for secondary and extraneous cues are likely to be useful in a wide range of tests" (1971, p. 365).

Operationally, TW can be defined as: "the ability to manifest test-taking skills which utilize the characteristics and formats of a test and/or test-taking situation in order to receive a score commensurate with the abilities being measured" (Oakland and Weilert, 1971). Ebel and Damrin (1960) treated TW as a specific cognitive skill, capable of being developed through experience. They considered TW as one of

the four "bases" from which examinees could respond to objective test questions, clearly separating this ability from the other three--direct knowledge of content; response sets; and chance guessing.

The basic issue involved in TW seems to be one of determining the extent to which a test validly discriminates on only those variables it was designed to measure (Oakland and Weilert, 1971). This recent statement is not at odds with the opinions expressed by most writers in this area (e.g. Thorndike, 1949; Ebel and Damrin, 1960; Vernon, 1962; Ebel, 1965; Millman and Setijadi, 1966). Several of these writers expressed the opinion that, on well-constructed tests, a lack of test-taking sophistication could be a large source of error in measurement. Rather than viewing TW as insignificant or undesirable, the consensus seems to be that tests should be constructed with greater care and that people should be given training in how to take tests.

Based on a review of several studies, Millman, Bishop and Ebel (1965) outlined the TW principles, grouping them as either dependent on or independent of the test constructor or purpose. In summarizing the state of TW research, they concluded: "There appears to be no systematic study of either the importance of test-wiseness or the degree to which it can be taught or measured" (1965, p. 707). The stated purpose of their analysis was to provide a framework within which future investigators could work, and they posed a series of questions for study. In spite of their excellent outline, very few studies since have focused directly on the problem. The terminology and framework they provided have been increasingly adopted in the research that has been done, how-

ever, so that some "common ground" seems to have resulted from their efforts.

One of the questions posed by Millman and his colleagues was whether or not TW can be taught. A number of recent studies have been directed to this question (e.g. Gibb, 1964; Moore, Schutz and Baker, 1966; Moore, 1968; Wahlstrom and Boersma, 1969; Slakter, Koehler and Hampton, 1970; Oakland and Weilert, 1971). In terms of the variety of learning experiences that have been designed, these studies reflect a rather broad-based approach to providing instruction in TW or related skills. There were varying degrees of success reported in these studies, and almost always, there was a criterion problem. Most of the programs and tests were designed for elementary or secondary school students, and no relevant studies focusing on an adult, non-college population were found. Although several programs designed to coach adults in dealing with specific tests (e.g. Civil Service, Armed Forces, CLEP) have been marketed, even the "popular" writers have not dealt with instruction in general test-wiseness. On the basis of those studies dealing with other than adult populations, it would appear that an individual's level of TW can be increased through training. Little evidence about persistence or the extent to which TW will generalize has been found.

Another question posed by Millman, Bishop and Ebel was related to the correlates of TW. No studies reflecting a comprehensive investigation of the correlates of TW were found, but several have focused on

selected variables. For the most part, discussions of the correlates of TW have emphasized test anxiety, response sets, general mental ability, and risk-taking. The biographical variables receiving greatest attention have been sex and grade level (or age), largely because of the concentration of studies using elementary or secondary students.

The nature of the relationship between TW and anxiety has not been demonstrated. There is some evidence that familiarity with item types might lessen anxiety in a classroom situation, but whether or not this type of familiarity could be considered TW is debatable (Sassenrath, 1967). Although the idea that test sophistication and test anxiety are not compatible is generally accepted, empirical evidence is lacking. The importance of response sets for personality test scores has been well demonstrated in the literature (e.g. Cronbach, 1950; Bass, 1955; Couch and Keniston, 1960; Wevrick, 1962; Stricker, 1969). However, the concept is seen as relatively unimportant in multiple choice tests of achievement (Cronbach, 1950). In fact, the whole concept of TW appears to be different in personality and achievement tests.

Risk-taking (on objective examinations) appears to be fairly consistent within a given test, but the relationship between this and TW remains to be demonstrated (Stone, 1962; Slakter, 1967). Slakter (1969) has suggested that a certain level of TW is essential before a subject can profit from taking risks. Although the feeling among researchers seems to be that general mental ability and TW are positively correlated (e.g. Stanley, 1971), little real proof of this has been offered. In at

least one study, the relationship between TW and general intelligence was not significant (Kreit, 1967). There is a similar paucity of research into the relationship of selected biographical characteristics to TW. Age has been shown to be positively correlated with TW for preschool through high school students. No data on the relationship between TW and age, or recency of test-taking experience, were available for adults.

It seems apparent that considerably more research into the nature of TW is needed. On the basis of a review of the recent literature, there would seem to be some agreement that people who are test-wise perform at a high level consistently, almost regardless of the type of test. There is evidence, however, that instructions in how to respond to specific types of items helps specifically. Stricker (1969) sees TW not as a broad, general ability, but rather as consisting of a set of "distinct and largely unrelated skills." Ebel and Damrin concluded that "insofar as 'test-taking' is a specific cognitive skill, it can, like any cognitive skill, be developed through experience. To the extent that differences in this skill are eliminated by adequate training, obtained differences in test scores will provide better estimates of true difference between the capacities and abilities of individuals" (1960, p. 1511).

The Problem

The CLU designation is awarded to qualified professionals in the insurance industry only upon successful completion of a series of ten achievement-type examinations. The examinations are prepared, administered

and evaluated by the American College of Life Underwriters, a non-profit organization which has been involved in this examination process for almost 50 years. In addition to examination preparation, the College prepares a variety of study guides and learning aids to assist candidates in attaining the CLU designation. Several other adult education programs also are offered. In all, the College currently serves approximately 60,000 students, administering examinations twice a year, in January and June.

The present study was initiated in response to a feeling among CLU candidates that they "understood the subject matter, but just couldn't pass the tests." This expression was in accord with a feeling among test developers and research staff at the College that the examination scores were probably contaminated somewhat by this population's lack of recent examination experience. This appeared a logical conclusion on the basis of the distributions of age and educational background of the CLU candidates. Approximately 35% of the candidates are 35 years of age or older when they begin their studies, and most have been away from an academic setting for quite a few years. It is entirely possible that a sizable number of new candidates have not taken an examination since high school or college: in some cases, a 20 or 30 year interim.

Many insurance companies now recommend that their company officers have the CLU designation, and, since the only way to obtain the designation is through successful completion of ten examinations, it would seem that this population would have a strong incentive to improve their test-taking abilities. Improvement in test-taking skills should in turn

improve the reliability and validity of the CLU examinations, decreasing the incidence of failure for reasons other than lack of knowledge.

For purposes of this study, TW was defined as a cognitive factor, one which is measurable and subject to change through either specific test experience or training in a test-taking strategy. Further, TW was assumed to be complex, related to certain personality characteristics, and in part specific to the nature of the test, the test situation and the examiner. The purpose was two-fold: to gather empirical evidence about the level of test-taking skills in the CLU population, and should the need exist, to develop an instructional program designed to improve these skills.

Test Development

In order to determine the level of TW in this population, it was necessary to construct a test to measure selected test-taking skills. Although some measures of TW had been developed as part of other studies, none were applicable to an adult population. The instrument developed for the measurement of TW consisted of 30 items, 10 items to measure each of three different TW skills. The test items were designed so that the correct answer could be determined only by blind guessing or by the application of a specific test-taking strategy.

Specifically, the test was designed to measure whether or not the examinee could arrive at the appropriate answer by: (1) recognizing and eliminating similar options; (2) recognizing and eliminating absurd options;

and (3) selecting an option which has a logical relationship with the stem. Skills 1 and 2 (referred to as "similar option" and "absurd option" skills), were included as deductive reasoning skills in the Millman, Bishop and Ebel classification, while skill 3 ("stem option"), was calssified as a cue using strategy (1965). The primary difference between these broad categories is that in the deductive reasoning strategies, it is not necessary to establish correlations between cues and the correct answer. These specific skills were selected because of the cognitive processes implicit in their utilization, and because they seemed to bear a close relationship to the types of skills which might be needed on the CLU examinations. Further, it was possible to assess the ability to apply these strategies directly in a test situation.

The TW items were all written by the author, then submitted to five judges for a content validity check. The judges were asked to sort the items into four stacks--one for each of the three TW skills with the fourth for items judged as not clearly reflecting any one of the skills. Items were retained only when there was unanimous agreement among the judges as to the nature of the TW skill measured. The items were pretested on two adult populations, revised and the test then compiled. Sample items are shown in Appendix A.

The 30 TW items were imbedded in a test consisting of 30 legitimate, general knowledge test items. The legitimate items, reflecting several content areas and utilizing item format similar to the TW items, were pretested on the same two adult populations. Only legitimate items of

difficulty levels from 50% to 90% and with discrimination in the appropriate direction were retained for use in the final form of the test. The decision to imbed the TW items within a set of legitimate items was made to avoid the possibly debilitating effects of the examinees' either "giving up" or feeling overly threatened during the examination. Since the TW items were not content based, the examinees would have very little, if any, positive reinforcement during the examination. It was hoped that some immediate positive reinforcement could be provided through the addition of legitimate items of fairly low difficulty. All test items were multiple choice and written in the formats commonly used for vocabulary, arithmetic calculations and general knowledge type items. The items were organized within the test according to subject matter and except for the arithmetic calculations items which were all legitimate, TW and legitimate items were ordered randomly within the test sections.

Survey of TW and CLU Population

To ascertain the level of test-taking proficiency in the CLU population, the TW Scale was administered to a total of 259 students enrolled in the "Individual Life and Health Insurance" course. A total of 15 different classes were tested. Class size ranged from 8 to 40, with a median of 15. Because the TW Scale is somewhat transparent if the purpose is known, its purpose was not disclosed to the students. All tests were administered in the regular classroom, during a class session by a trained test administrator.

Information related to each class, as well as biographical information about each student, was collected during the test administration. The descriptive statistics for each of the 15 classes are shown in Appendix B. Class means on the TW Scale ranged from 15.54 to 21.87, with individual scores ranging from 5 to 29. Reliability was estimated for the total test and for each of the subtests. The Cronbach alphas are shown in Figure I.

FIGURE I

TEST-WISENESS SCALE RELIABILITY

Subtest	Test Strategy	Number of Items	Alpha
Subtest I	Similar Option	10	0.44
Subtest II	Absurd Option	10	0.52
Subtest III	Stem Option	10	0.63
Total Test		30	0.73

In terms of the variation in TW level among this sample, the results of the first testing were considered sufficient to justify program development.

Test-Wiseness Program Development

The initial program was devoted exclusively to training in responding to objective, multiple-choice test items. The present version has been expanded to include instructions in responding to essay questions as well. In addition to these two sections, there is an Introduction primarily aimed at reduction of test-related anxiety. The program combines instruction and measurement in a workbook format, with some diagnostic testing and prescribed branching built-in. The Objective Examinations section is divided into three parts: an overview, in which the test-taking strategies are reviewed with examples; a diagnostic-branching section, requiring application of key strategies and providing specific instruction to program users as needed; and a final review test, sampling knowledge of principles and providing page references for review of questions answered incorrectly.

It is, therefore, a self-contained package of instruction, measurement and suggestions for review. Unlike most programs designed to teach test-taking, this program does not focus on practice in the types of items used in the CLU examinations. Although similar items were used to illustrate some of the principles, the focus was on instruction in specific strategies rather than on coaching in specific item types. A total of eleven strategies were included in the Objective Examination Section of the program. A listing of these strategies by level of treatment is shown in Figure II.

FIGURE II

STRATEGIES INCLUDED IN TW PROGRAM	
Knowledge Level (only)	Knowledge & Application Level
Time Using Strategy	Stem-Option Strategy (Cue-Using)
Error Avoidance Strategy	Similar Option Strategy (Deductive Reasoning)
Guessing Strategy	Absurd Option Strategy (Deductive Reasoning)
Conflicting Options Strategy (Deductive Reasoning)	Specific Determiners (Cue-Using)
Utilization of Relevant Content from other Items or Options (Deductive Reasoning)	
Grammatical Cues (Cue-Using)	
Intent Consideration Strategy	

Since all strategies are introduced at the knowledge level in the Overview of the Objective Examinations Section, this section does provide good coverage of all TW concepts. Application of selected strategies in a test situation is required in the diagnostic-branching section.

The diagnostic section has been organized as a series of separate instructional units, one for each of the four TW strategies taught at this level. The program user is to proceed through the strategies in a linear fashion, starting with the Absurd Option Strategy. Each instructional unit begins with two questions designed to determine whether or not the program user can apply the strategy in question. Depending upon his responses, he is either routed <u>through</u> the instructional unit, or around it, to the questions related to the next strategy. Each unit is completed via exit questions which also require application of the

strategy in question. At this point, the program user either reaches criterion and proceeds to the next strategy, or is routed back for another pass through the instructional unit.

Following the satisfactory completion of either the questions or instructional units for all strategies, the program user proceeds to the Final Review, a test designed to measure knowledge of the principles introduced within the program. Page references are provided for a review of questions answered incorrectly.

The Essay Examination Unit includes a brief introduction to the format of essay questions, some suggestions for setting a time schedule for test completion, and some general suggestions related to appearance of the essay and care in reading test directions. Some definitions of directives typically used in essay examinations also are included. It does not contain a diagnostic section, but does provide practice in applying the concepts introduced. Major emphasis is placed on discussions of organization of the essay and style of responding. Illustrative examples of acceptable responses and practice have been provided at strategic points. As the final practice, program users are asked to write an essay describing a picture. As with the Objective Examinations Unit, the Final Review covers all principles at the knowledge level and prescribes corrective review where needed.

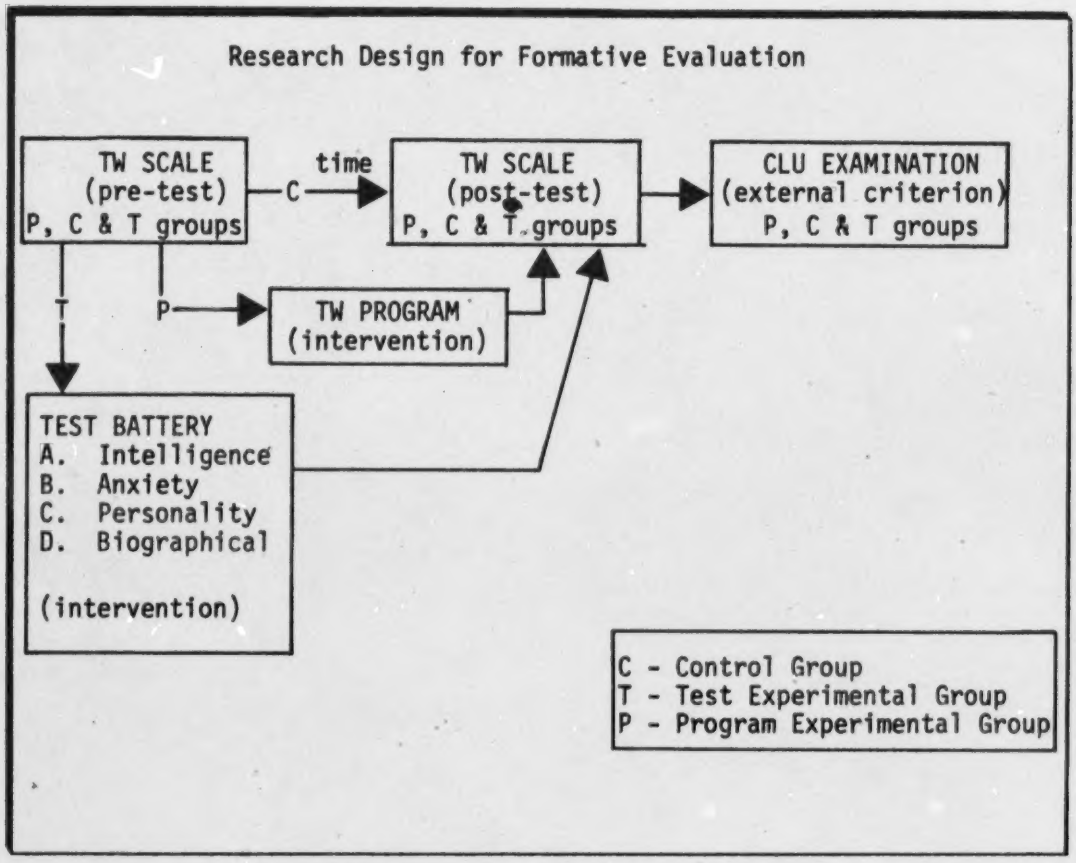
The complete program requires about four hours of working time to complete. To offer the flexibility of selecting specific sections or topics for study and review on an as-needed basis, the program has been packaged in a semi-module format. A special marking crayon and

latent ink printing greatly facilitate the recording of responses and provide branching instructions throughout the program.

Formative Evaluation

As part of the survey of TW in the CLU population, the TW Scale was administered to 259 CLU students enrolled in 15 "Individual Life and Health Insurance" classes. This served as the pre-test for the formative evaluation of the TW training program. Classes were assigned to one of three experimental groupings on the basis of matching class profiles on the biographical information collected during test administration. Class averages for these variables are included in Appendix B. As a result of matching, entire classes were assigned to one of the three groups: Program Experimental, Test Experimental, and Control, with five classes in each group. Original N's were 87, 92, and 80, respectively. A description of the research design is shown in Figure III on the following page.

FIGURE III



For the Program Experimental Group, the Objective Examinations section of the TW program was used as the intervention. A battery of psycological tests was administered to the Test Experimental Group shortly before the end of the semester. There was no intervention for the Control Group. The TW Scale described earlier served as the pre and post-test, and the same standardized CLU Examination was used as the external criterion for all groups.

It was hypothesized that the pre-test to post-test gain scores would be higher for the Program group than for the other two groups.

Further, it was hypothesized that classes in the Program group would have higher mean scores on the CLU examination.

Results and Implications

The Test Experimental group had been included in the design to provide information for a related study, and also to provide some information about the effects of recent systematic and comprehensive testing on the level of TW. Unfortunately, participation had to be on an individual volunteer basis and sample attrition was too high to interpret the results with meaning.

In comparing the pre- to post- gain scores for Program and Control groups, there appeared to be a greater gain in TW scores by the Program group. There were significant differences in the TW Total score for both groups, but even when the Control group gain was subtracted from the gain for the Program group, the difference remained significant. (Program group gain significant $>.001$; Control group gain significant $>.05$; Difference significant $>.05$). In addition, the Program Group showed a significant gain on the Similar Option subtest which was not shown by the Control Group (Sig. $>.001$). The evaluation of mean differences across groups substantiated this result. Comparing the TW Total score for the two groups resulted in a nonsignificant F for the pre-test, while the post-test showed a significant difference ($F = 6.24$; Sig. $>.001$). The comparison of Program and Control groups in terms of mean score on the CLU Examination resulted in no significant difference.

The findings show that training in TW produced the desired effect on an internal criterion, but no on an external criterion. The high

sample attrition must be considered with reference to both positive and negative findings. (Final N's: Program, 21; Control, 38). While the results of this phase of research are promising, subsequent research must incorporate some control for sample attrition, base rate performance (predicted performance), and maturation.

Based on an analysis of student-completed questionnaires, post-test information, and the completed programs, the original version of the TW program was revised. The revised program was expanded to include the information related to essay questions. This program was tested on a small sample (N=26) of students preparing for a mixed essay and objective CLU examination on a self-study basis. Because of the small N, a single group pre-test post-test design was used. All Ss took the same TW Scale as pre and post-tests, and completed the revised program. The results were interpreted exclusively for revision purposes.

Revisions were incorporated, as indicated, and the final program printed for distribution during the Spring, 1974, semester. A summative evaluation using a sample of Fall, 1974, CLU students has been completed and the results are included in another paper (Bajtelsmit, 1975).

A by-product of the research and development project was a contingency model approach to categorizing the test-taking principles. This model serves to provide meaningful distinctions among the various principles, rules and strategies. A rough model is attached as Appendix C; a more sophisticated interpretation is forthcoming.

APPENDIX A

Sample Items from the TW Scale

Similar Option Subtest:

1. Debilbous means

   A. Evil
   B. Wicked
   C. Bad
   D. Mercenary

2. The Roman philosopher Tuscus believed that

   A. The fall of Rome was inevitable.
   B. Nothing could be done to save the Roman empire.
   C. The directions of decay and decline of Rome were irreversible.
   D. Man, properly aware of the impending fall, could prevent it.

* * * * * *

Absurd Option Subtest:

1. Anexogas, a contemporary of Socrates is credited with writing

   A. "A History of Roman Civilization"
   B. A book criticizing the writings of Aristotle.
   C. A treatise on Greek democracy.
   D. The Dead Sea Scrolls.

2. Aspils are removed from certain flower seeds before they are planted in order to

   A. Keep birds from eating the flowers.
   B. Make the seeds germinate faster.
   C. Keep the flowers from blooming.
   D. Retain moisture.

* * * * * *

Stem Option Subtest:

1. Who is credited with founding the Messusan school of music?

   A. Bach
   B. Wagner
   C. Messuse
   D. Handel

2. The disease, intestio-phylitis, involves what part of the body?

   A. Lower back
   B. Intestinal tract
   C. Lungs and kidneys
   D. Stomach wall

APPENDIX B

TW SURVEY

CLU CLASS STATISTICS

CLASS	SIZE (N)	MEDIAN AGE (Years)	EDUCATION		YEARS SINCE LAST EXAM (Mean)	CLU EXAMS TAKEN		YRS. IN INSURANCE (Mean)	TEST-WISENESS PRE-TEST STATISTICS		
			H.S. (N)	COLL. (N)		ONE (N)	MORE THAN ONE (N)		Range	Mean	S.D.
1	21	32	9	3	2.1	1	7	6.4	15 to 28 (13)	21.05	2.84
2	11	35	7	2	3.1	0	1	4.1	15 to 27 (12)	20.27	3.53
3	28	33.5	13	4	3.6	1	2	5.3	13 to 28 (15)	18.79	3.54
4	15	27	1	9	2.3	0	0	4.2	18 to 28 (10)	21.87	2.68
5	17	31	3	10	2.4	4	0	6.4	12 to 26 (14)	20.6	4.34
6	9	34	2	5	1.8	0	4	7.4	14 to 27 (13)	20.22	3.89
7	17	36	10	3	2.6	0	3	6.3	5 to 29 (24)	17.47	5.78
8	8	33.5		2	2.5	1	1	4.4	13 to 28 (15)	19.38	4.8
9	13	39	7	3	3.1	0	0	8.4	11 to 25 (14)	18.31	3.74
10	40	30	8	14	3.7	3	0	5.7	13 to 27 (14)	18.85	3.72
11	32	31.5	9	10	2.8	0	1	3.7	10 to 29 (19)	18.5	4.3
12	11	26	4	4	2.4	0	1	5.8	7 to 22 (15)	15.54	4.97
13	10	34	8	1	2.9	2	3	7.8	14 to 26 (12)	19.5	3.72
14	13	37	2	4	2.8	1	3	8	15 to 26 (11)	20.38	3.63
15	14	27.5	4	3	2.1	1	2	4.9	16 to 27 (11)	20.5	3.31

APPENDIX C

TEST-WISENESS PRINCIPLES: A CONTINGENCY MODEL*

	GENERAL TEST-TAKING PRINCIPLES	PRINCIPLES SPECIFIC TO ITEM TYPE OR TEST CONSTRUCTOR
CONTENT DEPENDENT	1. Consider the subject matter and difficulty of neighboring items when interpreting and answering a given item. 2. Utilize relevant content information in other test items (options). 3. Consider the relevance of specific detail when answering a given item.	1. Eliminate options known to be incorrect and choose from among the remaining options. 2. Choose neither or both of two options which imply the correctness of each other. 3. Choose neither or one (but not both) of two statements, one of which, if correct, would imply the incorrectness of the other. 4. Restrict choice to those options which encompass two or more given statements known to be correct. 5. In an essay test, organize ideas logically.
CONTENT INDEPENDENT	1. Error-Avoidance strategies (care in reading and following directions, care in marking responses, check all answers). 2. Time-Using strategies (allocate time between items of equal worth equally, set up a schedule, etc.). 3. Guessing strategies (To be established once guessing penalties of test are known).	1. Make use of constructor's known tendencies: a) The correct option is longer (shorter) than the others, b) The correct option occupies a certain physical or logical position within the other options. 2. Make use of superfluous cues (not necessarily known as regular practices of the test constructor): a) Grammatical inconsistency between stem and all or some of the incorrect options, b) Inclusion of specific determiners, c) Resemblances between the option and an aspect of the stem.

*For illustrative purposes only, does not include all the Test-wiseness principles.

REFERENCES

Bass, B.M. "Authoritarianism or Acquiesence?" Journal of Abnormal and Social Psychology, 1955, 51, 616-623.

Cronbach, L.J. "Further Evidence on Response Sets and Test Design," Educational and Psychological Measurement, 1950, 10, 3-31.

Couch, A. & Keniston, K. "Yeasayers and Naysayers: Agreeing Response Sets as a Personality Variable," Journal of Abnormal and Social Psychology, 1960, 60, 151-174.

Ebel, R.L. Measuring Educational Achievement. New Jersey: Prentice-Hall, 1965.

Ebel, R.L. & Damrin, D.E. "Tests and Examinations." In Chester W Harris (ed.), Encyclopedia of Educational Research (3rd. ed.), New York: The Macmillan Company, 1960, 1502-1517.

Gibb, B.G. "Test-Wiseness as Secondary Cue Response," Unpublished doctoral dissertation, Stanford University, 1964.

Kreit, L. "The Effects of Test-Taking Practice on Pupil Test Performance," Unpublished doctoral dissertation, Indiana University, 1967.

Millman, J. & Setijadi. "A Comparison of the Performance of American and Indonesian Students on Three Types of Test Items," The Journal of Educational Research, 1966, 59, 273-275.

Millman, J., Bishop, C.H. & Ebel, R. "An Analysis of Test-Wiseness," Educational and Psychological Measurement, 1965, 25, 707-726.

Moore, J.C. "Manipulating the Effectiveness of a Self-Instructional Program," Journal of Educational Psychology, 1968, 59, 315-319.

Moore, J.C., Schutz, R.E. & Baker, R.L. "The Application of a Self-Instructional Technique to Develop a Test-Taking Strategy," American Educational Research Journal, 1966, 3, 13-17.

Oakland, T. & Weilert, E. "The Effects of Test-Wiseness Materials on Standardized Test Performance of Pre-School Disadvantaged Children," Paper presented at Feb. 1971 convention of The American Educational Research Association, New York.

Sassenrath, J.M. "Anxiety, Aptitude, Attitude and Achievement." Psychology in the School, 1967, 4, 341-346.

Slakter, M.J. "Generality of Risk Taking on Objective Examinations," Educational and Psychological Measurement, 1969, 29, 115-128.

Slakter, M.J. "Risk Taking on Objective Examinations." American Educational Research Journal, 1967, 4, 31-43.

Slakter, M.J., Koehler, R.A., Hampton, S.H. & Grennel, R.L. "Sex, Grade Level, and Risk Taking on Objective Examinations," Journal of Experimental Education, 1971, 39, 65-68.

Slakter, M.J., Koehler, R.A., & Hampton, S.H. "Grade Level, Sex, and Selected Aspects of Test-Wiseness," Journal of Educational Measurement, 1970a, 7, 119-122.

Slakter, M.J., Koehler, R.A., & Hampton, S.H. "Learning Test-Wiseness By Programmed Texts," Journal of Educational Measurement, 1970b, 7, 247-254.

Stanely, J.C. "Reliability," in Thorndike, R.L. (ed) Educational Measurement, Washington, D.C.: ACE, 1971, 356-442.

Stone, L.A. "Reliability of a Utility of Risk Measure," Psychological Reports, 1962, 10, 516.

Stricker, L.J. "Test-Wiseness on Personality Scales." Journal of Applied Psychology, 1969, 53, 1-18.

Thorndike, R.L. Personnel Selection: Test and Measurement Techniques. New York: John Wiley & Sons, 1949.

Vernon, P.E. "The Determinants of Reading Comprehension." Educational and Psychological Measurement, 1962, 22, 269-286.

Wahlstrom, M. & Boersma, F.J. "The Influence of Test-Wiseness Upon Achievement," Educational and Psychological Measurement, 1968, 28, 413-420.

Wevrick, L. "Response Set in a Multiple-Choice Test." Educational and Psychological Measurement, 1962, 22, 533-538.